


# Doubly uniparental inheritance and highly divergent mitochondrial genomes of the freshwater mussel *Unio tumidus* (Bivalvia: Unionidae)

Marianna Soroka · Artur Burzyński 

Received: 5 July 2016 / Revised: 12 December 2016 / Accepted: 28 January 2017 / Published online: 4 February 2017  
© The Author(s) 2017. This article is published with open access at Springerlink.com

**Abstract** *Unio tumidus* is a native European freshwater mussel from the family Unionidae. These mussels have a unique system of mitochondrial DNA inheritance called doubly uniparental inheritance (DUI). Under DUI, two types of mitochondrial DNA are present: haplotype F (female genome)—inherited from mother and haplotype M (male genome)—inherited from fathers to male offspring. The F genome occurs in eggs and in somatic tissues of both sexes, whereas the M genome is present in male gonads and gametes. We characterized three M and three F mitochondrial genomes of *Unio tumidus*. The lengths of these genomes were 15769–15770 bp for F type and 16607 bp for M type. In both genomes, the set of 38 genes which is typical for Unionidae was

identified: 37 metazoan genes and gender-specific ORFs. The non-coding sequences constituted only 5.2 and 3.5% of F and M genome, respectively. Both genomes were similarly high in average AT content (65–66%) but intraspecific nucleotide diversity amongst the three M genomes of *U. tumidus* was four times lower than amongst F genomes. The patterns of polymorphisms across the mitogenomes of the closest relatives confirmed that the M genomes accumulate more substitutions and the conserved regions within one lineage are also conserved in the other.

**Keywords** Complete mitochondrial genome · F and M mitochondrial genomes · DUI · Intra- and interspecific diversity

Guest editors: Manuel P. M. Lopes-Lima, Ronaldo G. Sousa, Lyuba E. Burlakova, Alexander Y. Karatayev & Knut Mehler / Ecology and Conservation of Freshwater Bivalves

**Electronic supplementary material** The online version of this article (doi:[10.1007/s10750-017-3113-7](https://doi.org/10.1007/s10750-017-3113-7)) contains supplementary material, which is available to authorized users.

M. Soroka (✉)  
Department of Genetics, Faculty of Biology, University of Szczecin, Felczaka 3c, 71-412 Szczecin, Poland  
e-mail: soroka.marianna@gmail.com

A. Burzyński  
Department of Genetics and Marine Biotechnology,  
Institute of Oceanology Polish Academy of Sciences,  
Powstańców Warszawy 55, 81-712 Sopot, Poland

## Introduction

Freshwater bivalves of the order Unionoida represent the largest bivalve radiation in freshwater. They are divided into six families (Etheriidae, Hyriidae, Iridinidae, Margaritiferidae, Mycetopodidae, and Unionidae), 181 genera, and about 840 species. The largest and most widespread of the six families of Unionoida is the Unionidae, with 674 species, that occur in all geographic regions (Graf & Cummings, 2007; Bogan, 2008; Bogan & Roe, 2008).

There are 14 species of Unionidae in Europe ([www.faunaeur.org](http://www.faunaeur.org), Araujo et al., 2005, 2009; Graf, 2007; Graf & Cummings, 2007), of which three are

considered endangered and are protected by law (*Unio crassus*, *Anodonta cygnea*, and *Pseudanodonta complanata*). Almost all European species are native, with the exception of *Sinanodonta woodiana*, originating from SouthEast Asia (Petró, 1984; Kiss, 1995; Afanasiev et al., 1997; Kraszewski & Zdanowski, 2001; Sousa et al., 2014; Labecka & Domagala, 2016). The species chosen for analysis in this study, *Unio tumidus*, is native to Central and Western Europe. It is relatively common in Poland, forming numerous and stable populations in rivers and lakes (Piechocki & Dyduch-Falniowska, 1993). Therefore, it is not expected to show genetic signs of recent rapid demographic processes such as bottleneck and founder effects.

In addition to biparentally inherited nuclear genome, a typical animal cell contains small, closed, circular DNA molecules in its mitochondria. These genomes are usually approximately 16 kb in size and very conservative with regard to gene contents and size. They contain 37 genes encoding 13 protein subunits of the enzymes of the oxidative phosphorylation (OXPHOS) system (*nad1-nad6*, *nad4L*, *cox1-cox3*, *cytb*, *atp6*, and *atp8*), the two rRNAs of the mitochondrial ribosome (*srRNA* and *lrRNA*), and the 22 tRNAs necessary for the translation of the proteins encoded by mtDNA (Avise, 1986; Moritz et al., 1987; Boore, 1999). These genomes usually exhibit strict maternal inheritance (SMI). However, a different mode of mitochondrial inheritance has been discovered in marine mussels *Mytilus* and called doubly uniparental inheritance (DUI) (Fisher & Skibinski, 1990).

Currently, DUI has been confirmed in several bivalve families (Mytilidae, Unionidae, Margaritiferidae, Hyriidae, Veneridae, Donacidae, Nuculanidae, Mactridae, Arctidae, and Solenidae), suggesting widespread occurrence of the phenomenon in bivalves (Skibinski et al., 1994; Zouros et al., 1994; Rawson & Hilbish, 1995; Hoeh et al., 1996a; Liu et al., 1996; Passamonti & Scali, 2001; Hoeh et al., 2002; Curole & Kocher, 2005; Walker et al., 2006; Soroka, 2008; Theologidis et al., 2008; Boyle & Etter, 2013; Huang et al., 2013; Plazzi, 2015; Soroka & Burzyński, 2015, 2016; Dégletagne et al., 2016). Under DUI, two types of mitochondrial DNA are present: haplotype F (female genome)—inherited according to SMI and haplotype M (male genome)—inherited from fathers to sons. Consequently, all males are

heteroplasmic for two, sometimes very divergent haplotypes. Usually, the M haplotype is predominantly present in male gonads and gametes, whereas in somatic tissues, the F haplotype dominates. Both M and F genomes are usually quite similar in size and gene contents but the M haplotype accumulates substitutions faster (Stewart et al., 1995; Hoeh et al., 1996b; Zouros, 2000; Zbawicka et al., 2010). Given the fact that mitochondrial genomes already evolve 5–10 times faster than typical single-copy nuclear DNA (Avise, 1986; Avise et al., 1987; Moritz et al., 1987), these genomes are examples of very dynamic genetic systems.

Gender-specific ORFs (open reading frames) have been discovered in M and F genomes of Unionidae (Breton et al., 2009; Breton et al., 2011; Ghiselli et al., 2013; Milani et al., 2013). These ORFs could be responsible for the different modes of transmission of the mtDNA and/or gender-specific adaptive functions of the M and F mtDNA genomes in unionoid bivalves (Breton et al., 2009). Despite the lack of any recognizable similarity to genes of known function and generally very fast evolution leading to rapid loss of sequence homology, these ORFs were identified in all published mt genomes of DUI Unionidae (Soroka, 2010a; Soroka & Burzyński, 2010; Huang et al., 2013; Mitchell et al., 2016).

In addition to gender-specific ORFs, the *cox2* gene differs substantially in length between F and M genomes of Unionidae: the M-type protein is usually more than 150 amino acids longer (Curole & Kocher, 2002, 2005). It has been postulated that it may be involved in functions other than electron transfer in complex IV, such as gamete maturation, fertilization, and/or embryogenesis (Chakrabarti et al., 2006, 2007, 2009; Chapman et al., 2008). These speculations are substantiated by earlier reports of enhanced expression of *cox2* during rat and human spermatogenesis (Saunders et al., 1993; Liang et al., 2004).

The aim of this study was to characterize M and F mitochondrial genomes from the European freshwater mussel *Unio tumidus*. The presence of DUI in this species was confirmed but only individual mitochondrial genes have been characterized (Soroka, 2010b). Here we present, sequenced for the first time, the complete sequences of both genomes, first such case in the genus *Unio* and second among European Unionidae after *Anodonta anatina* (Soroka & Burzyński,

2015, 2016). The motivation to increase the taxonomic coverage of M and F mitochondrial genomes stems from the belief that comparison between the two genomes should improve our understanding of mitochondrial genome evolution as this unusual system offers in essence the mitochondrial evolution in two replicates. Therefore, the focus of our comparative analysis is on the common parts of the two genomes.

## Materials and methods

Sampling of *U. tumidus* mussels was performed in northern Poland in May and June 2006, during the breeding season. Four mussels were collected from the river Oder near Szczecin (specimens 155, 157, 158, and 159) and two from the lake Sosnowe near Czluchów (specimens 203 and 207). The mussels were sexed by microscopic examination of gonads: specimens 157, 159, and 203 were males; specimens 155, 158 and 207 were females.

DNA was isolated using Qiagen DNeasy Tissue Kit (Germany). Mitochondrial DNA was amplified in two overlapping Long Range PCRs (LR-PCR), with lineage specific primers, as described previously (Soroka & Burzynski, 2010). The sequence of each LR-PCR amplicon was obtained by primer-walking approach (Zbawicka et al., 2007; Soroka & Burzynski, 2010). All PCR products were purified by ExoSap procedure (Werle et al., 1994) and sequenced directly using BigDye Terminator chemistry in Macrogen (South Korea).

Raw sequence reads were assembled in complete genomes using software tools from Staden package (Staden et al., 2001). Protein coding and rRNA annotations were conducted as described previously (Soroka & Burzyński, 2010). The prediction and annotation of tRNA gene positions and structures were done using *arwen* software (Laslett & Canbäck, 2008). The six complete mtDNA sequences of *U. tumidus* mitochondrial genomes sequenced in this study have been deposited in GenBank (KY021073–KY021078). All the information, including sex and specimen numbers, are included in GenBank records.

Selection of mitochondrial genomes for comparative analysis was done based on *Blast* searches against *nr* GenBank database (Camacho et al., 2008). Three criteria were used: availability of complete M and F genomes for the species, reasonably close genetic

distance to the sequences described here, and the same genome organization. Six species were selected for the final analysis (Table 1). Since the M and F genomes of Unionidae differ in structure, in order to allow their alignment, the longest possible colinear fragments of the genomes from both lineages were selected (Fig. 4; Table 1). Notably, the five species from the Gonideini (*Unio japonensis*, *Hyriopsis cumingii*, *Lamprotula leai*, *Potomida littoralis*, and *Solenia carinatus*) have also different order of genes between *cox2* and *srRNA* in the F genome (Breton et al., 2009) and were not considered for this analysis. The first fragment (A) spans the region of the genome starting at the *trnQ* gene an ending at *nad4L*. The second fragment (B) starts at *atp6* gene and ends at *nad3*; the third one (C) starts at *nad2* and ends at *nad5*. The alignment of the three fragments was constructed using clustalW algorithm (Thompson et al., 2002), for all seven species and both lineages, for a total of three alignments, 14 sequences each. One representative genome of *U. tumidus* was used in sliding window analysis but the final results did not depend on the choice. The obtained alignments were inspected for the potential problems requiring manual intervention or filtering of poorly aligned fragments. However, no such problems were encountered: all annotated features aligned well and the intergenic regions present in A, B, and C genomic fragments were very short. The single, long indel constituted the well known M *cox2* extension, and this part also aligned well, forming the expected pattern with a large insert in the M-type sequences. Therefore, in order to avoid data loss or bias, no filtering or trimming procedure was applied, taking also into account that sites with alignment gaps were excluded from all analyses anyway.

To assess and compare the patterns of polymorphisms across the genomes in intra and inter lineage comparisons, a sliding window approach was used in DnaSP (Librado & Rozas, 2009). The alignments of three colinear mitogenome fragments (Table 1) of the seven closest relatives were used for this purpose. The sequence polymorphism (expressed as nucleotide diversity,  $\pi$ ) within the F and M lineage was calculated in the sliding window of 200 bp along the alignment. Simultaneously, the genetic distance ( $D_{xy}$ ) between M and F groups was calculated. Alignment gaps were excluded from calculating diversity but did count towards the position of the midpoint of the sliding window. To recover the data for the area of *cox2*

**Table 1** Reference data for the Unionidae mitogenomes used in comparative analysis

| Species                            | Type | Accession | A           | B          | C           |
|------------------------------------|------|-----------|-------------|------------|-------------|
| <i>Pyganodon grandis</i>           | F    | FJ809754  | 10531–14020 | 14320–2615 | 3388–10086  |
|                                    | M    | FJ809755  | 10681–14123 | 15535–3219 | 3795–10502  |
| <i>Quadrula quadrula</i>           | F    | FJ809750  | 10678–14189 | 14475–2661 | 3547–10331  |
|                                    | M    | FJ809751  | 11119–14664 | 15336–3179 | 3559–10446  |
| <i>Venustaconcha ellipsiformis</i> | F    | FJ809753  | 10590–14126 | 14413–2676 | 3503–10281  |
|                                    | M    | FJ809752  | 10719–14446 | 15604–3197 | 3639–10411  |
| <i>Utterbackia peninsularis</i>    | F    | HM856636  | 10287–13772 | 14072–2617 | 3374–10033  |
|                                    | M    | HM856635  | 10579–14058 | 15209–3217 | 3683–10380  |
| <i>Lamprotula tortuosa</i>         | F    | KC109779  | 10427–13924 | 14197–2641 | 3447–10185  |
|                                    | M    | KC441487  | 10726–14277 | 15217–3215 | 3784–10576  |
| <i>Anodonta anatina</i>            | F    | KF030964  | 18–3496     | 3790–7933  | 8705–15413  |
|                                    | M    | KF030963  | 77–3840     | 4789–9521  | 10123–16836 |
| <i>Unio tumidus</i>                | F    | KY021077  | 27–3521     | 3804–7962  | 8764–15489  |
|                                    | M    | KY021073  | 79–3560     | 4538–9245  | 9717–16537  |

A, B, and C columns give the coordinates of the colinear fragments used in sliding window analysis

Complete genome references: Doucet-Beaupré et al. (2010), Breton et al. (2011), Wang et al. (2013) and Soroka & Burzynski (2015, 2016)

extension present in M genomes only, separate sliding window analysis was run within the M genome dataset.

Phylogenetic analysis was done using single nucleotide alignment obtained by concatenation of the three individual alignments (Table 1). Rational data partitioning was not possible due to several frameshift errors in the reference GenBank records; moreover, the phylogenetic signal was strong and consistent enough to use non-partitioned data safely. All six *U. tumidus* mitogenomic sequences were used in this analysis. The divergence of M and F mitogenomes greatly pre-dates the subfamily level differentiation in the studied group. This not only makes the M and F sequences the best choices of outgroups for each other but also poses difficulties when more distant sequences are aligned with them as more columns are then excluded from the analysis, lowering the resolution. Since the only purpose of the phylogenetic analysis was verification of the phylogenetic context of the newly sequenced mitogenomes, no outgroup sequences were used in comparative analyses. Both simple but computationally efficient Neighbour-Joining (NJ) phenetic method and more sophisticated Maximum Likelihood (ML) phylogenetic approach were used. The models of molecular evolution for ML reconstruction were evaluated using four criteria: hierarchical likelihood ratio test (hLRT), Bayesian information criterion (BIC), Akaike Information Criterion (AIC) as well as Akaike-corrected information criterion (AICc). The same GTR+G

model was favoured by all criteria (Supplementary Table S1) and used in subsequent phylogenetic reconstruction, using PHYML algorithm (Guindon & Gascuel, 2003). Rate heterogeneity among sites was approximated with four rate categories but all model parameters were estimated along with the tree topology. To evaluate stability of the topology, the bootstrap procedure was used with 100 replicates. All phylogenetic analyses were performed in CLC Genomics Workbench (QIAGEN). There was full congruence of the tree topologies obtained by both methods; therefore, only the results of ML method are presented.

The MEGA6 program (Tamura et al., 2013) was used to evaluate nucleotide composition and codon counts and relative synonymous codon usage (RSCU) (Sharp et al., 1986) as well as genetic distances:  $K_a$ ,  $K_s$  and protein  $p$ -distances, under default parameters. For the two amino acids having more than one mitochondrial tRNA (leucine and serine) RSCU was calculated separately for each set of codons. To calculate intraspecific nucleotide diversities uncorrected nucleotide  $p$ -distances were used. Free energy calculations ( $\Delta G$ ) and drawings of predicted tRNA structures were done using Vienna software (Lorenz et al., 2011), under the default set of parameters, except for the temperature. Rather than the default 37°C, the more realistic 20°C was used.

Genetic maps were created in CLC Genomics Workbench, except for the compositional bias charts which were made in *cgview* (Stothard & Wishart, 2005) and overlaid with the maps.





between *nad5* and *trnQ*; in the M genome between *trnH* and *trnQ*, NCR $\beta$  (between *trnF* and *nad5*), and NCR $\gamma$  (for the F genome between *nad3* and *trnA* excluding *trnH*; for the M genome between *nad3* and *trnA*). Together they constitute 59 and 52% of all NCRs in F and M genomes, respectively. Intralineage structural polymorphism in NCRs is negligible; there is only one indel in F NCR $\alpha$ . The M-F divergence in NCR $\alpha$ ,  $\beta$  and  $\gamma$  is quite high: 10, 19, and 32%, respectively, with the average at 26% (Table 2).

The NCRs are very AT rich (75%), even when the overall genome composition is taken into account. The AT content in both types of genomes is similar: 66.3% in the F genome and 65% in the M genomes. The compositional asymmetry (the differences in AT-skew) is associated with the direction of transcription (Fig. 1): lower-than-average AT-skew is present in parts transcribed in forward direction, while higher-than-average AT-skew is present in parts transcribed in reverse direction. This is also similar in M and F type genomes.

Codon usage was analysed across both types of genomes (Supplementary Table S2). Most codons showed visible bias, with RSCU values frequently substantially higher or lower than 1. However, the bias was very similar in M and F type genomes. Local compositional bias is the natural source of non-random use of codons in mitochondrial genomes. Selecting compositionally biased regions would be a subjective decision; therefore, codon usage bias was also analysed separately for genes differing in the direction of transcription as these features are correlated (Fig. 1), and the span of genes is far less subjective. An interesting pattern was found and is presented in Table 3. The two groups of genes showed consistent and very specific bias across both types of genomes. For all genes encoded on the “plus” strand, the codon having an “A” at the last position is always preferentially used. If the codons do not differ in the “A” contents, there was no pronounced bias. Contrary to this, for genes encoded on the “minus” strand, only the codons ending with “T” were always preferred.

#### Intraspecific genetic diversity

The intralineage nucleotide diversity ( $\pi$ ) among the three F genomes is 0.4%. There are 96 polymorphic sites, with majority of substitutions represented by transitions. The majority (77%) of polymorphic sites is

localized in CDS, but a significant fraction (15%) is also present in NCRs. The nucleotide diversity among the three M genomes is 0.1%; there are 20 polymorphic sites, 90% of them in CDS. However, all recorded substitutions in the F lineage are synonymous, while all substitutions recorded in the M lineage are non-synonymous (Table 2).

The average interlineage nucleotide diversity can only be measured in specific parts as the M and F genomes are not colinear. The distances recorded for individual genes of the M and F genomes are listed in Table 2. For example, the *rRNA* genes were only marginally longer in the M than in the F genomes, and there was no intraspecific intralineage length polymorphism. There were also no substitutions in the M lineage, whereas the F lineage polymorphism was low but measurable in *rRNA* genes (0.1 and 0.2% for *srRNA* and *lrRNA*, respectively). The genetic distance between the M and F genomes was relatively low, at 22% (Table 2).

Intralineage diversity in some tRNA genes was limited to at most single substitutions per gene. There were only six such cases. These substitutions did not influence the stability of tRNA structures (Fig. 2). The biggest change in free energy occurred in *trnG* from the F genome. In this case, the G/A substitution occurred within the stem structure. The mutated “A” version had substantially higher free energy ( $\Delta G = -10.02$ ) than the more frequent G variant, but the structure remained more stable than some other *trn* genes (for example in *trnC* from the M genome). The overall length of all *trn* genes in F genome is 1421 bp, with average length of a single gene at 65 bp. In the M genome, these values are a little higher, at 1448 and 66 bp, respectively. These parts of the studied genomes are very conserved, with the average distance between M and F genomes ranging from 14% in *trnG* to 35% in *trnS1* (24% average). Despite these differences all tRNAs encoded by both types of genomes use a “U” at the first anticodon position whenever possible, with the exception of only the start methionine tRNA which has classic CAU anticodon (instead of possible UAU).

#### Patterns of substitutions across the genomes

We wanted to focus our analysis on the common parts of both types of genomes as they essentially represent two replicates of mitochondrial genome in the same

**Table 2** Length of genes and genetic distances between F and M genomes of *U. tumidus*

| Gene                                      | Type | Length (bp) | <i>p</i> -D nt | <i>K<sub>s</sub></i> | <i>K<sub>a</sub></i> | <i>K<sub>a</sub>/K<sub>s</sub></i> | <i>p</i> -D aa |
|---|------|-------------|----------------|----------------------|----------------------|------------------------------------|----------------|
| Non-coding                                |      |             |                |                      |                      |                                    |                |
| NCR $\alpha$                              | F    | 306         | 0.009 (0.004)  | NA                   | NA                   | NA                                 | NA             |
|   | M    | 82          | 0              |                      |                      |                                    |                |
|   | F/M  |             | 0.102 (0.033)  |                      |                      |                                    |                |
| NCR $\beta$                               | F    | 32          | 0              | NA                   | NA                   | NA                                 | NA             |
|   | M    | 43          | 0              |                      |                      |                                    |                |
|   | F/M  |             | 0.194 (0.074)  |                      |                      |                                    |                |
| NCR $\gamma$                              | F    | 148         | 0.009 (0.006)  | NA                   | NA                   | NA                                 | NA             |
|   | M    | 183         | 0.004 (0.004)  |                      |                      |                                    |                |
|   | F/M  |             | 0.324 (0.038)  |                      |                      |                                    |                |
| All NCR                                   | F    | 827         | 0.011 (0.003)  | NA                   | NA                   | NA                                 | NA             |
|   | M    | 588         | 0.001 (0.001)  |                      |                      |                                    |                |
|   | F/M  |             | 0.259 (0.017)  |                      |                      |                                    |                |
| tRNA (only variable tRNA genes are shown) |      |             |                |                      |                      |                                    |                |
| tRNA-Ala                                  | F    | 65          | 0.011 (0.011)  | NA                   | NA                   | NA                                 | NA             |
|   | M    | 64          | 0              |                      |                      |                                    |                |
|   | F/M  |             | 0.301 (0.056)  |                      |                      |                                    |                |
| tRNA-Arg                                  | F    | 64          | 0.011 (0.011)  | NA                   | NA                   | NA                                 | NA             |
|   | M    | 67          | 0              |                      |                      |                                    |                |
|   | F/M  |             | 0.169 (0.048)  |                      |                      |                                    |                |
| tRNA-Cys                                  | F    | 62          | 0              | NA                   | NA                   | NA                                 | NA             |
|   | M    | 64          | 0.010 (0.010)  |                      |                      |                                    |                |
|   | F/M  |             | 0.180 (0.049)  |                      |                      |                                    |                |
| tRNA-Gly                                  | F    | 62          | 0.011 (0.011)  | NA                   | NA                   | NA                                 | NA             |
|   | M    | 62          | 0              |                      |                      |                                    |                |
|   | F/M  |             | 0.141 (0.044)  |                      |                      |                                    |                |
| tRNA-Met                                  | F    | 66          | 0.010 (0.010)  | NA                   | NA                   | NA                                 | NA             |
|   | M    | 62          | 0              |                      |                      |                                    |                |
|   | F/M  |             | 0.164 (0.047)  |                      |                      |                                    |                |
| tRNA-Thr                                  | F    | 66          | 0.010 (0.010)  | NA                   | NA                   | NA                                 | NA             |
|   | M    | 62          | 0              |                      |                      |                                    |                |
|   | F/M  |             | 0.219 (0.054)  |                      |                      |                                    |                |
| All tRNA                                  | F    | 1421        | 0.003 (0.001)  | NA                   | NA                   | NA                                 | NA             |
|   | M    | 1448        | 0              |                      |                      |                                    |                |
|   | F/M  |             | 0.240 (0.010)  |                      |                      |                                    |                |
| rRNA                                      |      |             |                |                      |                      |                                    |                |
| <i>srRNA</i>                              | F    | 859         | 0.001 (0.001)  | NA                   | NA                   | NA                                 | NA             |
|   | M    | 861         | 0              |                      |                      |                                    |                |
|   | F/M  |             | 0.210 (0.014)  |                      |                      |                                    |                |
| <i>lrRNA</i>                              | F    | 1278        | 0.002 (0.001)  | NA                   | NA                   | NA                                 | NA             |
|   | M    | 1315        | 0              |                      |                      |                                    |                |
|   | F/M  |             | 0.222 (0.012)  |                      |                      |                                    |                |
| All rRNA                                  | F    | 2137        | 0.002 (0.001)  | NA                   | NA                   | NA                                 | NA             |
|   | M    | 2176        | 0              |                      |                      |                                    |                |
|   | F/M  |             | 0.220 (0.009)  |                      |                      |                                    |                |

**Table 2** continued

| Gene         | Type | Length (bp) | <i>p</i> -D nt | $K_s$ | $K_a$ | $K_a/K_s$ | <i>p</i> -D aa |
|--------------|------|-------------|----------------|-------|-------|-----------|----------------|
| Protein      |      |             |                |       |       |           |                |
| <i>nad1</i>  | F    | 900         | 0.007 (0.002)  | 0.016 | 0.004 | 0.250     | 0.007 (0.004)  |
|              | M    | 906         | 0.001 (0.001)  | 0     | 0.001 | 0         | 0.002 (0.002)  |
|              | F/M  |             | 0.425 (0.015)  | 0.889 | 0.530 | 0.592     | 0.478 (0.028)  |
| <i>nad6</i>  | F    | 489         | 0.012 (0.004)  | 0.005 | 0.016 | 3.2       | 0.029 (0.010)  |
|              | M    | 540         | 0.001 (0.001)  | 0.004 | 0     | 0         | 0.000 (0.000)  |
|              | F/M  |             | 0.565 (0.022)  | 0.950 | 1.096 | 1.154     | 0.704 (0.035)  |
| <i>nad4</i>  | F    | 1347        | 0.004 (0.001)  | 0.010 | 0.001 | 0.100     | 0.003 (0.002)  |
|              | M    | 1329        | 0.003 (0.001)  | 0.005 | 0.001 | 0.200     | 0.003 (0.002)  |
|              | F/M  |             | 0.443 (0.014)  | 0.757 | 0.631 | 0.834     | 0.504 (0.023)  |
| <i>nad4L</i> | F    | 297         | 0              | 0     | 0     | 0         | 0              |
|              | M    | 267         | 0.003 (0.002)  | 0.009 | 0     | 0         | 0              |
|              | F/M  |             | 0.456 (0.030)  | 0.672 | 0.736 | 1.100     | 0.605 (0.052)  |
| <i>atp6</i>  | F    | 702         | 0.006 (0.002)  | 0.016 | 0.001 | 0.625     | 0.003 (0.003)  |
|              | M    | 693         | 0.001 (0.001)  | 0.003 | 0     | 0         | 0              |
|              | F/M  |             | 0.424 (0.018)  | 0.623 | 0.626 | 1         | 0.534 (0.033)  |
| <i>atp8</i>  | F    | 195         | 0              | 0     | 0     | 0         | 0              |
|              | M    | 180         | 0              | 0     | 0     | 0         | 0              |
|              | F/M  |             | 0.471 (0.037)  | 0.681 | 0.778 | 1.142     | 0.632 (0.064)  |
| <i>cox3</i>  | F    | 780         | 0.005 (0.002)  | 0.009 | 0.004 | 0.444     | 0.008 (0.004)  |
|              | M    | 780         | 0.002 (0.001)  | 0.006 | 0     | 0         | 0              |
|              | F/M  |             | 0.338 (0.016)  | 0.636 | 0.377 | 0.593     | 0.409 (0.030)  |
| <i>cox1</i>  | F    | 1542        | 0.001 (0.001)  | 0.004 | 0     | 0         | 0              |
|              | M    | 1599        | 0.001 (0.001)  | 0.001 | 0.001 | 1         | 0.001 (0.001)  |
|              | F/M  |             | 0.288 (0.012)  | 0.648 | 0.263 | 0.409     | 0.311 (0.020)  |
| <i>cox2</i>  | FM   | 681         | 0.004 (0.002)  | 0.007 | 0.003 | 0.429     | 0.006 (0.004)  |
|              | M    | 1272        | 0.001 (0.001)  | 0.003 | 0     | 0         | 0              |
|              | F/M  |             | 0.398 (0.018)  | 0.708 | 0.511 | 0.722     | 0.496 (0.033)  |
| <i>nad3</i>  | F    | 357         | 0.004 (0.003)  | 0.016 | 0     | 0         | 0              |
|              | M    | 351         | 0              | 0     | 0     | 0         | 0              |
|              | F/M  |             | 0.465 (0.028)  | 1.503 | 0.605 | 0.403     | 0.586 (0.036)  |
| <i>nad2</i>  | F    | 966         | 0.003 (0.001)  | 0.009 | 0.001 | 0.111     | 0.002 (0.002)  |
|              | M    | 981         | 0.001 (0.001)  | 0.003 | 0     | 0         | 0              |
|              | F/M  |             | 0.511 (0.015)  | 2.125 | 0.686 | 0.323     | 0.628 (0.026)  |
| <i>cytb</i>  | F    | 1149        | 0.005 (0.002)  | 0.012 | 0.003 | 0.250     | 0.005 (0.003)  |
|              | M    | 1152        | 0.001 (0.001)  | 0.002 | 0     | 0         | 0              |
|              | F/M  |             | 0.365 (0.014)  | 0.859 | 0.380 | 0.442     | 0.405 (0.024)  |
| <i>nad5</i>  | F    | 1734        | 0.004 (0.001)  | 0.011 | 0.001 | 0.091     | 0.002 (0.002)  |
|              | M    | 1767        | 0.001 (0.001)  | 0     | 0.001 | 0         | 0.001 (0.001)  |
|              | F/M  |             | 0.441 (0.011)  | 0.701 | 0.649 | 0.926     | 0.560 (0.020)  |

The *p*-distance at nucleotide and amino acid level as well as synonymous ( $K_s$ ) and non-synonymous ( $K_a$ ) distances are shown.  $K_s$  and  $K_a$  were calculated using Nei–Gojobori method (Nei & Gojobori, 1986). Standard error estimates (in parentheses) were obtained by a bootstrap procedure (2,000 replicates). All positions containing gaps were eliminated. Analyses were performed in MEGA6 (Tamura et al., 2013)



**Table 3** Codon usage of genes encoded on “plus” and “minus” strands of mitochondrial genomes of *U. tumidus*

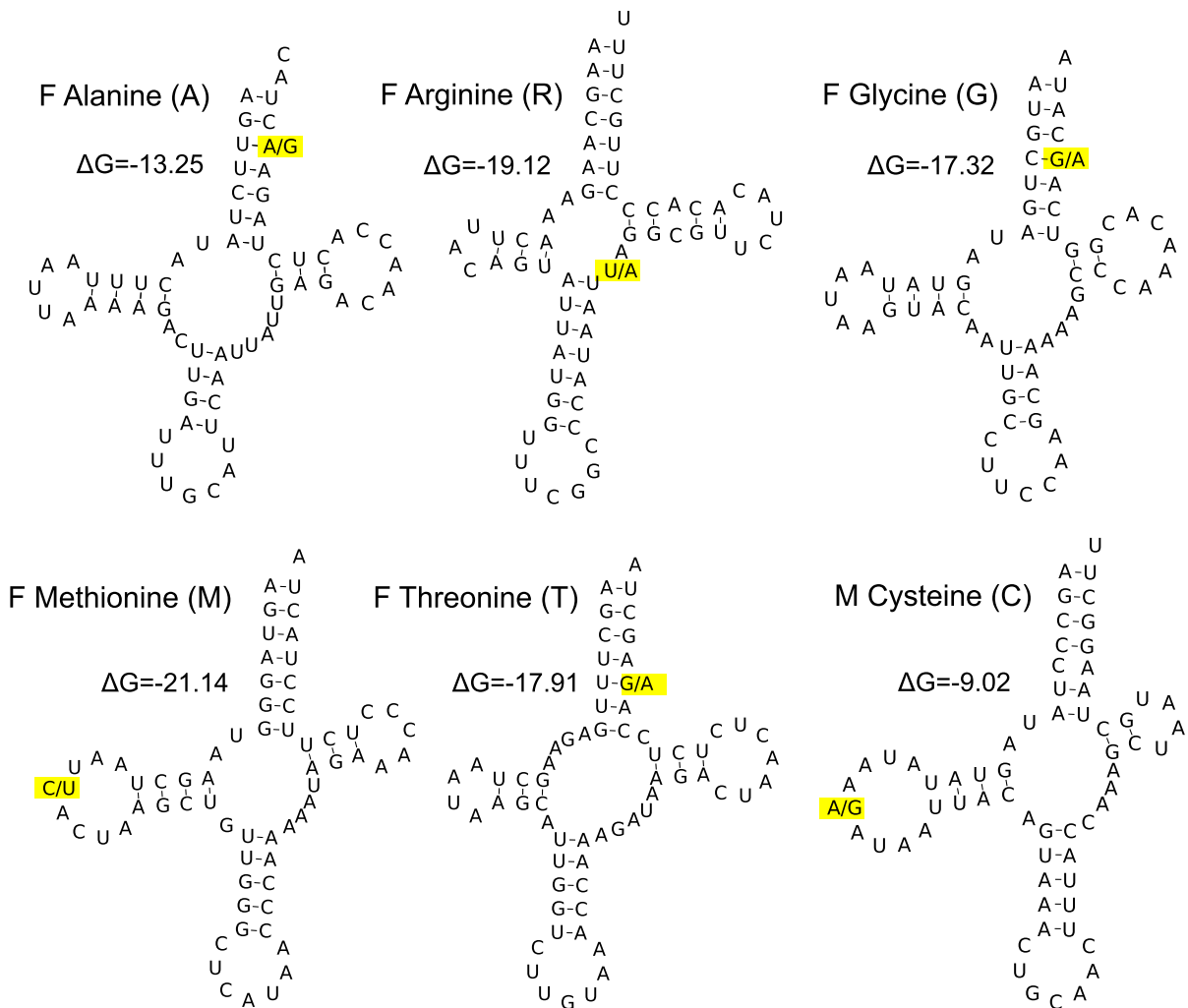
| Codon | aa | Plus  |       | Minus |       | Codon | aa | Plus  |       | Minus |       |
|-------|----|-------|-------|-------|-------|-------|----|-------|-------|-------|-------|
|       |    | Count | RSCU  | Count | RSCU  |       |    | Count | RSCU  | Count | RSCU  |
| TAA   | *  | 24    | 2.000 | 30    | 1.053 | AAC   | N  | 171   | 1.046 | 50    | 0.288 |
| TAG   | *  | 0     | 0.000 | 27    | 0.947 | AAT   | N  | 156   | 0.954 | 297   | 1.712 |
| GCA   | A  | 216   | 1.834 | 97    | 0.540 | CCA   | P  | 177   | 1.983 | 107   | 0.915 |
| GCC   | A  | 117   | 0.994 | 29    | 0.162 | CCC   | P  | 57    | 0.639 | 31    | 0.265 |
| GCG   | A  | 6     | 0.051 | 50    | 0.279 | CCG   | P  | 15    | 0.168 | 36    | 0.308 |
| GCT   | A  | 132   | 1.121 | 542   | 3.019 | CCT   | P  | 108   | 1.210 | 294   | 2.513 |
| TGC   | C  | 27    | 0.885 | 53    | 0.280 | CAA   | Q  | 151   | 1.924 | 138   | 1.195 |
| TGT   | C  | 34    | 1.115 | 325   | 1.720 | CAG   | Q  | 6     | 0.076 | 93    | 0.805 |
| GAC   | D  | 39    | 0.963 | 30    | 0.153 | CGA   | R  | 63    | 2.571 | 89    | 1.333 |
| GAT   | D  | 42    | 1.037 | 362   | 1.847 | CGC   | R  | 14    | 0.571 | 3     | 0.045 |
| GAA   | E  | 113   | 1.592 | 195   | 0.863 | CGG   | R  | 3     | 0.122 | 61    | 0.914 |
| GAG   | E  | 29    | 0.408 | 257   | 1.137 | CGT   | R  | 18    | 0.735 | 114   | 1.708 |
| TTC   | F  | 179   | 0.637 | 84    | 0.122 | AGA   | S  | 122   | 2.228 | 257   | 1.021 |
| TTT   | F  | 383   | 1.363 | 1293  | 1.878 | AGC   | S  | 43    | 0.785 | 63    | 0.250 |
| GGA   | G  | 154   | 2.000 | 406   | 1.081 | AGG   | S  | 12    | 0.219 | 340   | 1.351 |
| GGC   | G  | 51    | 0.662 | 74    | 0.197 | AGT   | S  | 42    | 0.767 | 347   | 1.378 |
| GGG   | G  | 47    | 0.610 | 488   | 1.300 | TCA   | S  | 188   | 1.617 | 159   | 0.822 |
| GGT   | G  | 56    | 0.727 | 534   | 1.422 | TCC   | S  | 131   | 1.127 | 45    | 0.233 |
| CAC   | H  | 87    | 1.381 | 35    | 0.206 | TCG   | S  | 18    | 0.155 | 29    | 0.150 |
| CAT   | H  | 39    | 0.619 | 304   | 1.794 | TCT   | S  | 128   | 1.101 | 541   | 2.796 |
| ATC   | I  | 208   | 0.643 | 42    | 0.087 | ACA   | T  | 226   | 1.671 | 102   | 0.761 |
| ATT   | I  | 439   | 1.357 | 925   | 1.913 | ACC   | T  | 140   | 1.035 | 36    | 0.269 |
| AAA   | K  | 213   | 1.760 | 210   | 0.854 | ACG   | T  | 6     | 0.044 | 63    | 0.470 |
| AAG   | K  | 29    | 0.240 | 282   | 1.146 | ACT   | T  | 169   | 1.250 | 335   | 2.500 |
| CTA   | L  | 335   | 1.985 | 101   | 0.805 | GTA   | V  | 119   | 1.488 | 473   | 0.979 |
| CTC   | L  | 124   | 0.735 | 24    | 0.191 | GTC   | V  | 60    | 0.750 | 70    | 0.145 |
| CTG   | L  | 27    | 0.160 | 46    | 0.367 | GTG   | V  | 27    | 0.338 | 319   | 0.660 |
| CTT   | L  | 189   | 1.120 | 331   | 2.637 | GTT   | V  | 114   | 1.425 | 1,070 | 2.215 |
| TTA   | L  | 520   | 1.825 | 1,037 | 1.136 | TGA   | W  | 117   | 1.614 | 278   | 1.035 |
| TTG   | L  | 50    | 0.175 | 788   | 0.864 | TGG   | W  | 28    | 0.386 | 259   | 0.965 |
| ATA   | M  | 401   | 1.681 | 465   | 1.150 | TAC   | Y  | 144   | 1.003 | 94    | 0.325 |
| ATG   | M  | 76    | 0.319 | 344   | 0.850 | TAT   | Y  | 143   | 0.997 | 485   | 1.675 |

Absolute counts of all CDS from six genomes are given as well as RSCU. The unbiased RSCU values should be around 1,000

phylogenetic context. The mitogenomic data from six available Unionidae M and F pairs along with six *U. tumidus* genomes were used in phylogenetic analysis. There was full congruence between two methods used (ML and NJ), both in terms of full support of the obtained topology and in similarity of relative branch lengths. The ML tree is presented in Fig. 3 as an example. Full congruence between topologies of M and F clades can be observed.

To access and compare the patterns of polymorphisms across the genomes in intra- and interlineage

comparisons, a sliding window approach was used (Fig. 4). The general trend of the M genomes to accumulate more substitutions is visible. On the other hand, the conserved regions within one lineage are also conserved in the other. The genes regarded as the most conserved parts of the genomes like for example *rRNA* genes with clusters of *tm* genes flanking them can be identified as conserved in each lineage separately as well as in interlineage comparison. However, the patterns of divergences are far from uniform, and some parts of the M genome are more



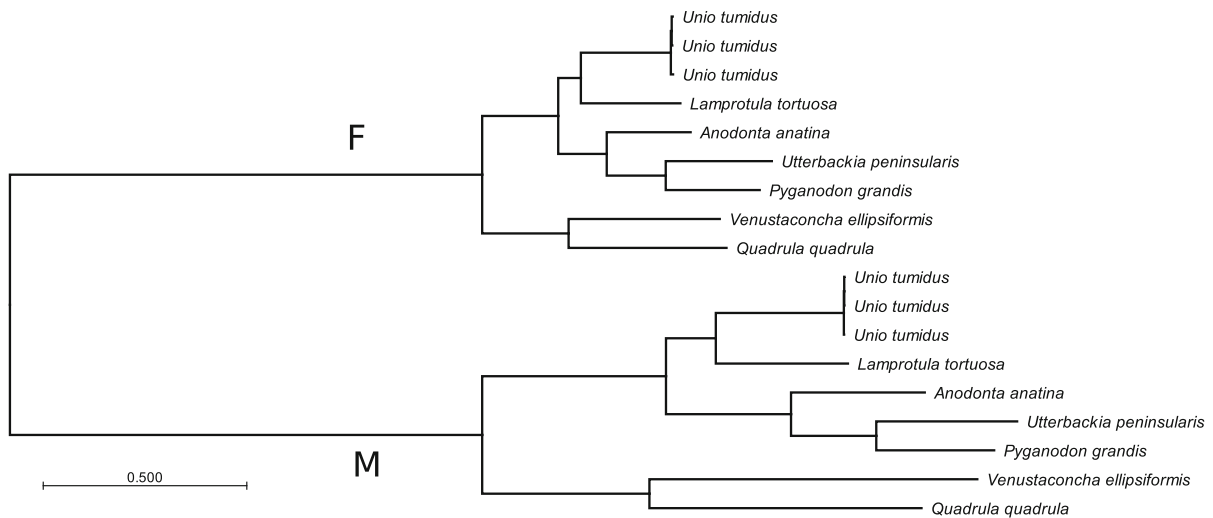
**Fig. 2** The secondary structures of polymorphic tRNA genes, along with their free energies. See text for details

variable than other parts of the F genome: middle part of *cox1* in M is less variable than the beginning of *atp6* in F, for example. There are regions of the F genome which are much more polymorphic than even the average M genome polymorphism (ea. the 3' flank of the *nad6* in A). However, all three lines are roughly parallel, indicating that the more polymorphic regions in M are usually also more polymorphic in F, and the same regions are also responsible for higher between lineage divergences. There are only few exceptions where the polymorphisms do not match. One is located in fragment B, within the *cox1* gene, in apparently most conservative parts of the genome. The other is located within the NCR adjacent to *nad5*, in this case also the overall pattern of M-F divergence is in conflict

with the apparently more conservative M lineage. The third, less obvious anomaly is at the beginning of the *cox2* gene, where the F lineage is apparently more conserved than expected based of M-F and M comparisons.

## Discussion

Typical set of 38 genes was identified in female and male mitochondrial genomes of *U. tumidus*, including both the canonical metazoan set of 37 genes and a gender-specific ORF characteristic for Unionidae (Avisé, 1986; Boore, 1999; Breton et al., 2009, 2011). The genetic distance between the two



**Fig. 3** Phylogenetic tree based on the concatenated alignment of large colinear fragments from reference M and F genomes mentioned in Table 1 and all *U. tumidus* genomes sequenced in this paper. There were 13520 positions in the final dataset.

types of sequenced genomes was higher than in most DUI bivalves, a feature typical for Unionidae (Bettinazzi et al., 2016). Gene order did not differ from the order reported for the majority of Unionidae F genomes, with 27 genes on the formally L strand and 11 on the H strand, although the difference of mass between the two strands is barely noticeable in this species, the H strand is only 1.1% heavier than the L strand.

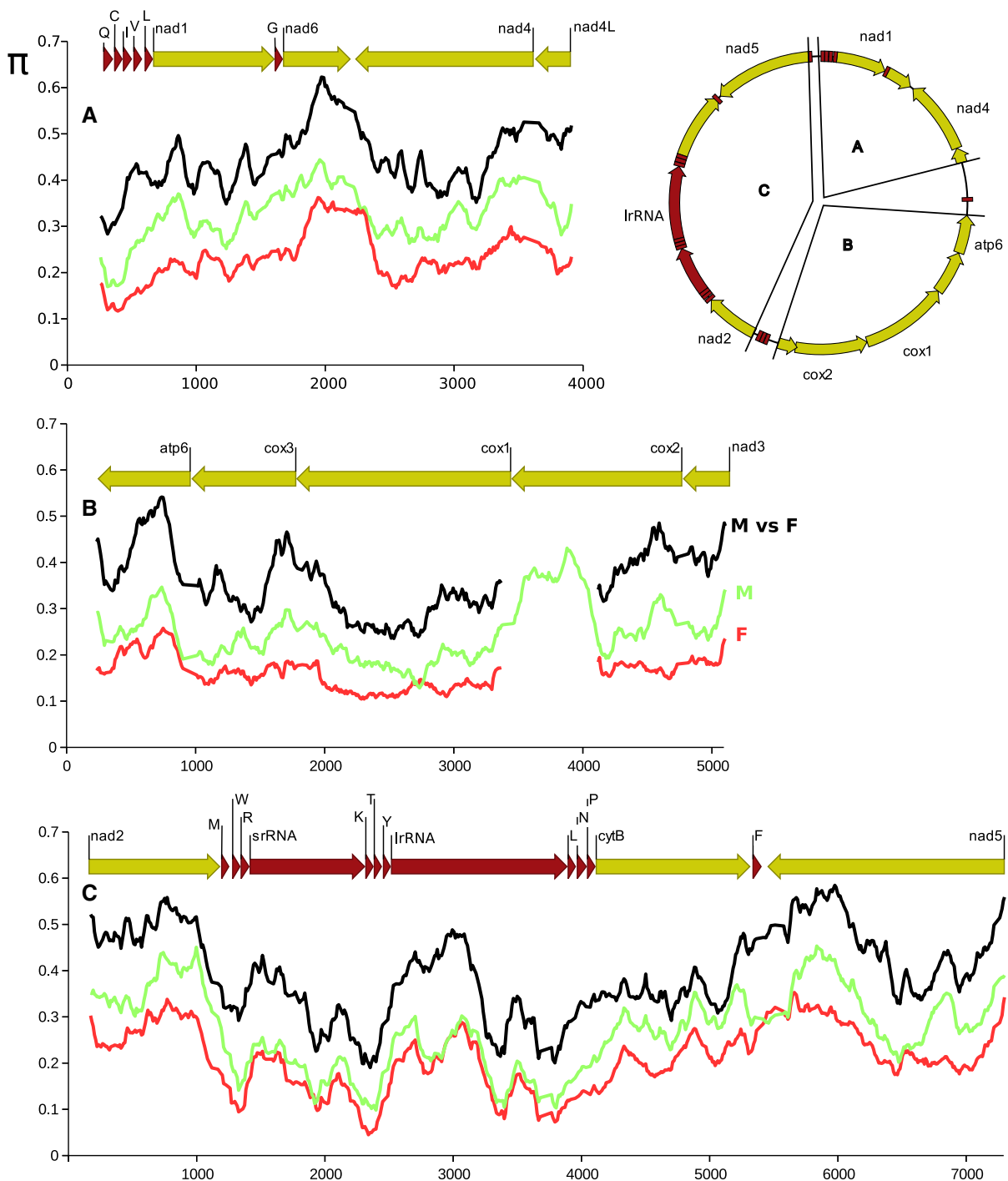
In Unionidae, M genomes are longer than F genomes (by about 7%), mainly due to the presence of the *cox2* extension (Curole & Kocher, 2005; Breton et al., 2007; Chakrabarti et al., 2007; Chapman et al., 2008; Huang et al., 2013). In *U. tumidus*, the difference between F and M genome is only 5.3%, similar to *Quadrula quadrula* (5.8%), but still much greater than the 2.3% reported in *Solenia carinatus* (Huang et al., 2013). This relatively smaller-than-average difference is not due to the shorter *cox2* extension; on the contrary, this extension is actually longer in *U. tumidus* (591 bp) than in other reported genomes (from 543 to 582 bp). However, the longer M genome contains smaller amount of NCRs (3.5% only), hence the difference.

The longest non-coding regions (NCRs) are supposed to play a key role in mtDNA functioning and maintenance. In Unionidae, the localization of these important functions is problematic because the NCRs

Maximum Likelihood method under GTR nucleotide substitution model with  $\Gamma$  distributed heterogeneity among sites was used. The bootstrap support for all bipartitions was 100%. The presented tree is arbitrary rooted at midpoint

are unusually short, and several candidates are considered. In *U. tumidus*, all the major NCRs reported for other Unionidae are present: NCR $\alpha$ , NCR $\beta$ , and NCR $\gamma$  (Breton et al., 2009). NCR $\beta$  in *U. tumidus* is the shortest (43 bp) but very AT rich (91 and 81%, respectively, for F and M). NCR $\alpha$  and NCR $\gamma$  are longer (82–306 bp) and moderately AT rich (74–76%), in line with the data for other Unionidae (Breton et al., 2009; Soroka, 2010a; Soroka & Burzyński, 2010; Plazzi et al., 2013). This is in sharp contrast with NCRs of marine bivalves having longer (>1100 bp) and less AT rich (60–66%) NCRs (Cao et al., 2004; Mizi et al., 2005; Burzyński & Śmietanka, 2009; Guerra et al., 2014).

Although the three sequences are not expected to be representative for the intraspecific diversity, the comparison hints at low overall intraspecific diversity. The values formally calculated from such data show low level of polymorphism, ranging from 0.03% in F genome of *U. pictorum* to 0.4% in the F genome of *U. tumidus* and 0.4% in *Anodonta anatina* (Soroka & Burzyński, 2015). These nucleotide diversities may indicate small effective population size of these species in Poland (Soroka & Burzyński, 2010). This is further confirmed by similar level of M genome polymorphism, at the level of 0.1% in two of the species, *U. tumidus* and *A. anatina* (Soroka & Burzyński, 2016). The overall sequence polymorphism observed in



**Fig. 4** Nucleotide diversity calculated in a sliding window of 200 bp in 10 bp steps along the three alignments of the fragments of the genome indicated by the top right panel. Seven species with both M and F lineages were used, as listed in Table 1. Red line indicates nucleotide diversity within the group

of seven F genomes, and green line indicates the diversity within the group of seven M genomes, while the black line indicates the distance between the two groups, expressed as Dxy. The calculations were done in DnaSP. The approximate positions of the genes in the alignments are shown above each plot

intraspecific comparisons is not surprising. However, the comparison of M and F data seems to contradict the general pattern of usually observed higher polymorphism in the M lineage: polymorphism observed in the F lineage is marginally higher in *U. tumidus*. However, since the values for both genomes are relatively low and the sample size is too small to perform statistical tests, these data can only be cautiously interpreted as unexpected.

The strong correlation between local AT-skew and the direction of translation has been noted previously (Soroka & Burzyński, 2010, 2015) and seems to be characteristic for all Unionidae mitochondrial genomes, with possible exception of MORF gene (Fig. 1). The apparent co-occurrence of a very strong codon usage bias seems to indicate common origin of both phenomena. However, the simplest explanation of the codon usage bias by compositional skew does not hold since the direction of the codon usage bias is not always correlated with the direction of the skew. All transcripts are enriched in T (negative AT-skew in forward and positive AT-skew in reverse direction). However, only the genes encoded in reverse use T-bias codons, as expected. The forward encoded genes use A-biased codons instead. There are three exceptions from the correlation between AT-skew value and the transcriptional direction in the M genome: the MORF, *nad3*, and the *cox2* extension. All these genes are known for their relatively relaxed selection hinting at the balance between selection and mutational pressure as the sources of these compositional phenomena.

The phylogenetic analysis is showing the usual gender-joining pattern typical of unionids (Fig. 3). However, the patterns of substitutions across the alignment are more informative. Remarkably, the lines in Fig. 4 are nearly parallel indicating the congruence between factors shaping the polymorphism of these genomes both within each lineage (M and F) and in the general phylogenetic context of the seven related species. Moreover, the deviations from this rule point at the regions of particular importance, the area between *trnF* and *nad5* indicates the greater importance of this NCR for the M than for the F genome, perhaps related to its function. RefSeq database contains sub-genomic annotations for all mitochondrial encoded gene products. According to this database, the part of the *cox1* showing similar pattern is responsible for the interaction with nuclear subunit of the OXPHOS complex, *cox6a*. Because it

seems to pose a very strong constrain in both lineages as well as between lineages, it is tempting to speculate that this gene exists in one version only in Unionidae. Given the well known difference in *cox2* between lineages, it is not surprising that this whole gene may experience different constrains in each lineage (Chapman et al., 2008).

**Acknowledgements** Financial support was provided by the Polish Ministry of Science and Higher Education through a Grant No. N 303 364 33 to MS.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Afanasiev, S., A. Shatokhina & B. Zdanowski, 1997. Some aspects of thermal tolerance of *Anodonta* from heated konińskie lake. Archives of Polish Fisheries 5: 5–11.
- Araujo, R., I. Gómez & A. Machordom, 2005. The identity and biology of *Unio mancus* Lamarck, 1819 (= *U. elongatulus*) (Bivalvia: Unionidae) in the Iberian Peninsula. Journal of Molluscan Studies 71: 25–31.
- Araujo, R., C. Toledo & A. Machordom, 2009. Redescription of *Unio gibbus* Spengler, 1793, a west palaearctic freshwater mussel with hookless glochidia. Malacologia 51: 131–141.
- Avise, J. C., 1986. Mitochondrial DNA and the evolutionary genetics of higher animals. Philosophical Transactions of the Royal Society of London B 312: 325–342.
- Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, A. C. Reeb & N. C. Saunders, 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics 18: 489–522.
- Bettinazzi, S., F. Plazzi & M. Passamonti, 2016. The complete female- and male-transmitted mitochondrial genome of *Meretrix lamarckii*. PLoS One 11: e0153631.
- Bogan, A. E., 2008. Global diversity of freshwater mussels (Mollusca, Bivalvia) in freshwater. Hydrobiologia 595: 139–147.
- Bogan, A. E. & K. J. Roe, 2008. Freshwater bivalve (Unioniformes) diversity, systematics, and evolution: status and future directions. Journal of the North American Benthological Society 27: 349–369.
- Boore, J. L., 1999. Animal mitochondrial genomes. Nucleic Acids Research 27: 1767–1780.
- Boyle, E. E. & R. J. Etter, 2013. Heteroplasmy in a deep-sea protobranch bivalve suggests an ancient origin of doubly uniparental inheritance of mitochondria in Bivalvia. Marine Biology 160: 413–422.

- Breton, S., H. D. Beaupré, D. T. Stewart, W. R. Hoeh & P. U. Blier, 2007. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? *Trends in Genetics* 23: 465–474.
- Breton, S., H. D. Beaupré, D. T. Stewart, H. Piontkivska, M. Karmakar, A. E. Bogan, P. U. Blier & W. R. Hoeh, 2009. Comparative mitochondrial genomics of freshwater mussels (Bivalvia: Unionoida) with doubly uniparental inheritance of mtDNA: gender-specific open reading frames and putative origins of replication. *Genetics* 183: 1575–1589.
- Breton, S., D. T. Stewart, S. Shepardson, R. J. Trdan, A. E. Bogan, E. G. Chapman, A. J. Ruminas, H. Piontkivska & W. R. Hoeh, 2011. Novel protein genes in animal mtDNA: a new sex determination system in freshwater mussels (Bivalvia: Unionoida)? *Molecular Biology and Evolution* 28: 1645–1659.
- Burzyński, A. & B. Śmietanka, 2009. Is interlineage recombination responsible for low divergence of mitochondrial *nad3* genes in *Mytilus galloprovincialis*? *Molecular Biology and Evolution* 26: 1441–1445.
- Camacho, C., G. Coulouris, V. Avagyan, et al., 2008. BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421.
- Cao, L., E. Kenchington, E. Zouros & G. C. Rodakis, 2004. Evidence that the large noncoding sequence is the main control region of maternally and paternally transmitted mitochondrial genomes of the marine mussel (*Mytilus* spp.). *Genetics* 176: 835–850.
- Chakrabarti, R., J. M. Walker, D. T. Stewart, R. J. Trdan, S. Vijayaraghavan, J. P. Curole & W. Hoeh, 2006. Presence of a unique male-specific extension of C-terminus to the cytochrome c oxidase subunit II protein coded by the male-transmitted mitochondrial genome of *Venustaconcha ellipsiformis* (Bivalvia: Unionoidea). *FEBS Letters* 580: 862–866.
- Chakrabarti, R., J. M. Walker, E. G. Chapman, S. P. Shepardson, R. J. Trdan, J. Curole, G. Watters, D. Stewart, S. Vijayaraghavan & W. Hoeh, 2007. Reproductive function for a C-terminus extended, male-transmitted cytochrome c oxidase subunit II protein expressed in both spermatozoa and eggs. *FEBS Letters* 581: 5213–5219.
- Chakrabarti, R., S. Shepardson, M. Karmakar, R. Trdan, J. Walker, R. Shandilya, D. Stewart, S. Vijayaraghavan & W. Hoeh, 2009. Extra-mitochondrial localization and likely reproductive function of a female-transmitted cytochrome c oxidase subunit II protein. *Development, Growth and Differentiation* 51: 511–519.
- Chapman, E. G., H. Piontkivska, J. M. Walker, D. T. Stewart, J. P. Curole & R. W. Hoeh, 2008. Extreme primary and secondary protein structure variability in the chimeric male-transmitted cytochrome c oxidase subunit II protein in freshwater mussels: evidence for an elevated amino acid substitution rate in the face of domain-specific purifying selection. *BMC Evolutionary Biology* 8: 165–181.
- Curole, J. P. & T. D. Kocher, 2002. Ancient sex-specific extension of the cytochrome c oxidase II gene in Bivalves and the fidelity of Double-Uniparental Inheritance. *Molecular Biology and Evolution* 19: 1323–1328.
- Curole, J. P. & T. D. Kocher, 2005. Evolution of a unique mitotype-specific protein-coding extension of the cytochrome c oxidase II gene in freshwater mussels (Bivalvia: Unionoida). *Journal of Molecular Evolution* 61: 381–389.
- Dégletagne, C., D. Abele & C. Held, 2016. A distinct mitochondrial genome with DUI-like inheritance in the ocean quahog *Arctica islandica*. *Molecular Biology and Evolution* 33: 375–383.
- Doucet-Beaupré, H., S. Breton, E. Chapman, P. Blier, A. Bogan, D. Stewart & W. Hoeh, 2010. Mitochondrial phylogenomics of the Bivalvia (Mollusca): searching for the origin and mitogenomic correlates of doubly uniparental inheritance of mtDNA. *BMC Evolutionary Biology* 10: 50.
- Fisher, C. & D. O. F. Skibinski, 1990. Sex-biased mitochondrial-DNA heteroplasmy in the marine mussel *Mytilus*. *Proceedings of the Royal Society of London Series B, Biological Sciences* 242: 149–156.
- Ghiselli, F., L. Milani, D. Guerra, P. L. Chang, S. Breton, S. V. Nuzhdin & M. Passamonti, 2013. Structure, transcription, and variability of metazoan mitochondrial genome: perspectives from an unusual mitochondrial inheritance system. *Genome Biology and Evolution* 5: 1535–1554.
- Graf, D. L., 2007. Palearctic freshwater mussel (Mollusca: Bivalvia: Unionoida) diversity and the Comparative Method as a species concept. *Proceeding of the Academy of Natural Sciences of Philadelphia* 156: 71–88.
- Graf, D. L. & K. S. Cummings, 2007. Review of the systematics and global diversity of freshwater mussel species (Bivalvia: Unionoida). *Journal of Molluscan Studies* 73: 291–314.
- Guerra, D., F. Ghiselli & M. Passamonti, 2014. The largest unassigned regions of the male- and female-transmitted mitochondrial DNAs in *Musculista senhousia* (Bivalvia: Mytilidae). *Gene* 536: 316–325.
- Guindon, S. & O. Gascuel, 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 52: 696–704.
- Hoeh, W. R., D. T. Stewart, B. W. Sutherland & E. Zouros, 1996a. Cytochrome c oxidase sequence comparisons suggest an unusually high rate of mitochondrial DNA evolution in *Mytilus* (Mollusca: Bivalvia). *Molecular Biology and Evolution* 13: 418–421.
- Hoeh, W. R., D. T. Stewart, B. W. Sutherland & E. Zouros, 1996b. Multiple origins of gender-associated mitochondrial DNA lineages in bivalves (Mollusca: Bivalvia). *Evolution* 50: 2276–2286.
- Hoeh, W. R., D. T. Stewart & S. I. Guttman, 2002. High fidelity of mitochondrial genome transmission under the doubly uniparental mode of inheritance in freshwater mussels (Bivalvia: Unionoidea). *Evolution* 56: 2252–2261.
- Huang, X., J. Rong, Y. Liu, M. Zhang, Y. Wan, S. Ouyang, C. Zhou & X. Wu, 2013. The complete maternally and paternally inherited mitochondrial genomes of the endangered freshwater mussel *Solenia carinatus* (Bivalvia: Unionidae) and implications for Unionidae taxonomy. *PLoS One* 8: e84352.
- Kiss, Á., 1995. The Propagation, Growth and Biomass of the Chinese Huge Mussel (*Anodonta woodiana woodiana* Lea, 1834) in Hungary. Univ. Agric Sci. Gödöllő, Tropical and Subtropical Department, Private Edition, 2nd edn: 1–33.



- Kraszewski, A. & B. Zdanowski, 2001. The distribution and abundance of the Chinese mussels *Anodonta woodiana* (Lea, 1834) in the heated Konin lakes. *Archives of Polish Fisheries* 9: 253–265.
- Labecka, A. M. & J. Domagala, 2016. Continuous reproduction of *Sinanodonta woodiana* (Lea, 1824) females – an invasive mussel species in a female-biased population. *Hydrobiologia*. doi:10.1007/s10750-016-2835-2.
- Laslett, D. & B. Canbäck, 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* 24: 172–175.
- Liang, G., X. D. Zhang, L. J. Wang, Y. S. Sha, J. C. Zhang, S. Y. Miao, S. D. Zong, L. F. Wang & S. S. Koide, 2004. Identification of differentially expressed genes of primary spermatocyte against round spermatid isolated from human testis using the laser capture microdissection technique. *Cell Research* 14: 507–512.
- Librado, P. & J. Rozas, 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
- Liu, H. P., J. B. Mitton & S. K. Wu, 1996. Paternal mitochondrial DNA differentiation far exceeds maternal mitochondrial DNA and allozyme differentiation in the fresh-water mussel, *Anodonta grandis grandis*. *Evolution* 50: 952–957.
- Lorenz, R., S. H. Bernhart, C. Höner zu Siederdisen, H. Tafer, C. Flamm, P. F. Stadler & I. L. Hofacker, 2011. ViennaRNA Package 2.0. *Algorithms for Molecular Biology* 6: 26.
- Milani, L., F. Ghiselli, D. Guerra, S. Breton & M. Passamonti, 2013. A comparative analysis of mitochondrial ORFans: new clues on their origin and role in species with Doubly Uniparental Inheritance of mitochondria. *Genome Biology and Evolution* 5: 1408–1434.
- Mitchell, A., D. Guerra, D. Stewart & S. Breton, 2016. In silico analyses of mitochondrial ORFans in freshwater mussels (Bivalvia: Unionoida) provide a framework for future studies of their origin and function. *BMC Genomics* 17. <http://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-016-2986-6>.
- Mizi, A., E. Zouros, N. Moschonas & G. C. Rodakis, 2005. The complete maternal and paternal mitochondrial genomes of the Mediterranean mussel *Mytilus galloprovincialis*: implications for the doubly uniparental inheritance mode of mtDNA. *Molecular Biology and Evolution* 22: 952–967.
- Moritz, C., T. Dowling & W. Brown, 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology, Evolution, and Systematics* 18: 269–292.
- Nei, M. & T. Gojobori, 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* 3: 418–426.
- Passamonti, M. & V. Scali, 2001. Gender-associated mitochondrial DNA heteroplasmy in the venerid clam *Tapes philippinarum* (Mollusca: Bivalvia). *Current Genetics* 39: 117–124.
- Petró, E., 1984. The occurrence of *Anodonta woodiana woodiana* in Hungary. *Állattani Közlemények* 84: 189–191.
- Piechocki, A. & A. Dyduch-Falniowska, 1993. Mięczaki (Mollusca). Małże (Bivalvia). Fauna Słownikowa Polski 7A, PWN, Warszawa.
- Plazzi, F., 2015. The detection of sex-linked heteroplasmy in *Pseudocardium sachalinense* (Bivalvia: Mactridae) and its implications for the distribution of doubly uniparental inheritance of mitochondrial DNA. *Journal of Zoological Systematics and Evolutionary Research* 53: 205–210.
- Plazzi, F., A. Ribani & M. Passamonti, 2013. The complete mitochondrial genome of *Solemya velum* (Mollusca: Bivalvia) and its relationships with Conchifera. *BMC Genomics* 14: 409.
- Rawson, P. D. & T. J. Hilbish, 1995. Evolutionary relationships among the male and female mitochondrial DNA lineages in the *Mytilus edulis* species complex. *Molecular Biology and Evolution* 12: 893–901.
- Saunders, P. T. K., M. R. Millar, A. P. West & R. M. Sharpe, 1993. Mitochondrial cytochrome c oxidase II messenger ribonucleic acid is expressed in pachytene spermatocytes at high levels and in a stage-dependent manner during spermatogenesis in the rat. *Biology of Reproduction* 48: 57–67.
- Sharp, P. M., T. M. Tuohy & K. R. Mosurski, 1986. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Research* 14: 5125–5143.
- Skibinski, D. O. F., C. Gallagher & C. M. Beynon, 1994. Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. *Genetics* 138: 801–809.
- Soroka, M., 2008. Doubly uniparental inheritance of mitochondrial DNA in the freshwater bivalve *Anodonta woodiana* (Bivalvia: Unionidae). *Folia Biologica (Krakow)* 56: 91–95.
- Soroka, M., 2010a. Characteristics of mitochondrial DNA of unionid bivalves (Mollusca: Bivalvia: Unionidae). II. Comparison of complete sequences of maternally inherited mitochondrial genomes of *Sinanodonta woodiana* and *Unio pictorum*. *Folia Malacologica* 18: 189–209.
- Soroka, M., 2010b. Characteristics of mitochondrial DNA of unionid bivalves (Mollusca: Bivalvia: Unionidae). I. Detection and characteristic of double uniparental inheritance (DUI) of unionid mitochondrial DNA. *Folia Malacologica* 18: 147–188.
- Soroka, M. & A. Burzyński, 2010. Complete sequences of maternally inherited mitochondrial genomes in mussels *Unio pictorum* (Bivalvia, Unionidae). *Journal of Applied Genetics* 51: 469–476.
- Soroka, M. & A. Burzyński, 2015. Complete female mitochondrial genome of *Anodonta anatina* (Mollusca: Unionidae): confirmation of a novel protein-coding gene (F ORF). *Mitochondrial DNA* 26: 267–269.
- Soroka, M. & A. Burzyński, 2016. Complete male mitochondrial genome of *Anodonta anatina* (Mollusca: Unionidae). *Mitochondrial DNA* 27: 1679–1680.
- Sousa, R., A. Novais, R. Costa & D. L. Strayer, 2014. Invasive bivalves in fresh waters: impacts from individuals to ecosystems and possible control strategies. *Hydrobiologia* 735: 233–251.
- Staden, R., D. P. Judge & J. K. Bonfield, 2001. Sequence assembly and finishing methods, 2nd ed. Wiley, New York.
- Stewart, D. T., C. Saavedra, R. R. Stanwood, A. O. Ball & E. Zouros, 1995. Male and female mitochondrial DNA lineages in the blue mussel (*Mytilus edulis*) species group. *Molecular Biology and Evolution* 12: 735–747.

- Stothard, P. & D. S. Wishart, 2005. Circular genome visualization and exploration using CGView. *Bioinformatics* 21: 537–539.
- Tamura, K., G. Stecher, D. Peterson, A. Filipski & S. Kumar, 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30: 2725–2729.
- Theologidis, I., S. Fodelianakis, M. B. Gaspar & E. Zouros, 2008. Doubly uniparental inheritance (DUI) of mitochondrial DNA in *Donax trunculus* (Bivalvia: Donacidae) and the problem of its sporadic detection in Bivalvia. *Evolution* 62: 959–970.
- Thompson, J. D., T. J. Gibson & D. G. Higgins, 2002. Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics* Chapter 2: Unit 2.3. doi:[10.1002/0471250953.bi0203s00](https://doi.org/10.1002/0471250953.bi0203s00).
- Walker, J. M., J. P. Curole, D. E. Wade, E. G. Chapman, A. E. Bogan, G. T. Watters & W. R. Hoeh, 2006. Taxonomic distribution and phylogenetic utility of gender-associated mitochondrial genomes in the Unionoida (Bivalvia). *Malacologia* 48: 265–282.
- Wang, G., X. Cao & J. Li, 2013. Complete F-type mitochondrial genome of Chinese freshwater mussel *Lamprotula tortuosa*. *Mitochondrial DNA* 24: 513–515.
- Werle, E., C. Schneider, M. Renner, M. Volker & W. Fiehn, 1994. Convenient single-step, one tube purification of PCR products for direct sequencing. *Nucleic Acids Research* 22: 4354–4355.
- Zbawicka, M., A. Burzyński & R. Wenne, 2007. Complete sequences of mitochondrial genomes from the Baltic mussel *Mytilus trossulus*. *Gene* 406: 191–198.
- Zbawicka, M., A. Burzyński, D. Skibinski & R. Wenne, 2010. Scottish *Mytilus trossulus* mussels retain ancestral mitochondrial DNA: complete sequences of male and female mtDNA genomes. *Gene* 456: 45–53.
- Zouros, E., 2000. The exceptional mitochondrial DNA system of the mussel family Mytilidae. *Genes and Genetic Systems* 75: 313–318.
- Zouros, E., A. O. Ball, C. Saavedra & K. R. Freeman, 1994. An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Proceeding of the National Academy of Sciences of United States of America* 91: 7463–7467.